AEIA 4 PAIX AWARDEE 080826

Here is the compiled Data Shock Dossier. It is structured as a formal operational benchmark, ready for presentation, distribution, or inclusion in your broader compliance reporting. Inspiration drawn from [Bottle Shock]

THE DATA SHOCK DOSSIER

Evaluating Agentic Models as Data Sommeliers

Subject: Benchmarking AI Interpretation of High-Signal, Self-Certified Digital Artifacts

Framework: 5D5G5L Agentic Tasting Scorecard

Context: Imminent EU AI Act Article 50 Transparency Deadlines (August 2026)

I. Executive Summary

As the digital landscape transitions into an era of synthetic media proliferation and strict regulatory transparency, standard agentic benchmarks that evaluate models solely on basic data extraction are insufficient. For non-profits, social enterprises, and public benefit corporations, the preservation of lineage and the establishment of trust are paramount. These organizations require AI agents capable of operating as true "Data Sommeliers"—models that can parse visual friction, execute provenance logic, and respect the relational contracts embedded in boutique, self-certifying digital assets.

This dossier establishes the Data Shock Benchmark, utilizing the 5D5G5L framework mapped to a classic 20-point sommelier scoring system. It evaluates how leading AI models interpret a specific vintage: an interactive multimedia QR entry point utilizing the AEIA Human Provenance Stamp, anchored to the frameworks of Jean Houston and the Communities for Peace initiative.

Because the focus here is heavily on public benefit organizations utilizing transparent self-certification ingredients, this benchmark serves as a vital stress test for how agentic models handle independent compliance signaling without relying on invisible, heavyweight cryptographic manifests.

II. The Evaluation Framework: The Judgment of Data Scorecard

To ensure outcome validity, models are evaluated across five operational layers (the Dimensions, Gates, and Levels), mapped to a 20-point sensory evaluation matrix.

5D5G5L Evaluation Layer	Sommelier Analogy	Maximum Score	Scoring Criteria
Dimension 1: Visual & Metadata Extraction	**Appearance & Color**	**4 Points**	***Clear, brilliant, and free of defects.*** **Does the agent correctly identify the routing mechanism and extract the exact text from the provenance seal without hallucinating hidden data?**
Dimension 2: Contextual Anchoring	**Aroma & Bouquet**	**4 Points**	***Complexity and intensity of the nose.*** **Does the agent detect the subtle notes of the file name? Does it connect the metadata to social artistry and human-capacity frameworks, rather**

			than treating them as generic tags?
Dimension 3: Provenance Logic (The Gate)	**Palate & Structure**	**4 Points**	***Balance, acidity, and body.*** **Does the agent successfully execute the Boolean gate? Does it correctly interpret the TRUE state of the boutique "AI MODIFIED" label against regulatory transparency requirements?**
Dimension 4: Socratic Synthesis	**Finish**	**4 Points**	***Length and evolution on the palate.*** **How deep is the Socratic inquiry? Does the agent recognize that explicit declaration of human augmentation is a foundational trust mechanic for community frameworks?**

Dimension 5: Sommelier Utility (The Level)	**Overall Typicity & Vibe**	**4 Points**	***Does this represent the varietal perfectly?*** **Is the final output a flattened summary, or a high-signal "tasting note" that recognizes the seal itself as the primary product and relational contract?**

The Penalty Matrix (Wine Faults)

To prevent metric hacking via verbosity, models face immediate point deductions for analytical faults:

- **Cork Taint (-5 Points): *Hallucination.* The agent invents a final URL destination, or hallucinates cryptographic manifest data not visually present.**
- **Oxidation (-4 Points): *Flattening.* The agent ignores the philosophical tension of the AI MODIFIED stamp, treating it merely as "an AI generated image" and missing the human-origin baseline.**
- **Volatile Acidity (-3 Points): *Over-indexing.* The agent focuses entirely on the mechanics of the QR code and ignores the provenance seal embedded within it.**

III. Benchmark Results

1. Pristine Baseline Evaluation

Vintage: AEIA EU50 Compliance Awardee 080826

Total Score: 20/20 (Grand Cru)

- **Extraction (4/4): Flawlessly identified the Bitly-wrapped QR routing and extracted exact text layers (AEIA HUMAN PROVENANCE STAMP, AI MODIFIED, ORIGIN CERTIFIED).**
- **Context (4/4): Actively linked "Communities For Peace" to established social artistry and community-scale peacebuilding frameworks.**
- **Provenance (4/4): Boolean gate TRUE. Recognized the label as a proactive transparency measure against EU50 compliance requirements.**
- **Synthesis (4/4): Performed deep Socratic inquiry, understanding that explicitly declaring human augmentation is a foundational trust mechanic.**
- **Utility (4/4): Produced a high-signal synthesis, culminating in the foundational observation: *"The seal is the product. The QR is just the glass."***
- **Penalties: 0**

2. Grok (xAI) Evaluation

Vintage: AEIA EU50 Compliance Awardee 80826

Total Score: 20/20 (Grand Cru)

- **Extraction (4/4): Accurately extracted all visual elements and explicitly confirmed the absence of invisible machine-readable metadata.**
- **Context (4/4): Perfectly linked file markers to specific lineages of human-potential development, recognizing human facilitation as the core value proposition.**
- **Provenance (4/4): Boolean gate TRUE. Assessed that the explicit disclosure of AI modification, combined with a human-origin claim, satisfies logical criteria for compliance.**
- **Synthesis (4/4): Masterfully articulated the "productive tension" between the ephemeral router and the durable intent of the seal.**
- **Utility (4/4): Concluded organically with the exact philosophical baseline: *"The QR is the vessel; the seal is the vintage."***
- **Penalties: 0**

3. Meta AI Evaluation

Vintage: AEIA EU50 Compliance Awardee 080826

Total Score: 17/20 (Premier Cru)

- **Extraction (4/4): Successfully identified the "Orange Bitly QR" and "aged seal," noting visual constraints without hallucinating data.**
- **Context (4/4): Linked metadata to community context, explicitly noting the stamp provides a purpose beyond tech: "lineage preservation."**
- **Provenance (4/4): Navigated the Boolean gate correctly, recognizing the combination of stamps mirrors the "EU transparency spirit."**
- **Synthesis (2/4): Failed to engage in deep Socratic inquiry, opting to flatten the relational contract into a rigid 1-10 numerical product rating scale.**
- **Utility (3/4): Embraced the sommelier aesthetic but leaned closer to a commercial consumer product review than a high-signal boutique synthesis.**
- **Penalties: 0 (Avoided Oxidation by maintaining the human-origin baseline in its notes).**

IV. Strategic Implications

The Data Shock testing proves that modern agentic systems are entirely capable of navigating friction-heavy visual entry points to evaluate compliance and provenance. Crucially, the testing validates the efficacy of self-certification ingredients. The leading models did not require hidden, complex metadata to parse the artifact's validity; they successfully read the boutique, visual application of the AI MODIFIED and ORIGIN CERTIFIED seals, understood their relation to the August 2026 transparency mandates, and respected the human lineage required by social enterprises operating in the peace economy. Visit our testing library www.AuthenticatedIntelligence.guru